doi:10.1016/j.jmb.2011.07.063 J. Mol. Biol. (2011) 413, 51–65

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Degenerate RNA Packaging Signals in the Genome of Satellite Tobacco Necrosis Virus: Implications for the Assembly of a *T*=1 Capsid

David H. J. Bunka [1]†, Stephen W. Lane [1]†, Claire L. Lane [1], Eric C. Dykeman [2], Robert J. Ford [1], Amy M. Barker [1], Reidun Twarock [2], Simon E. V. Phillips [1,3] and Peter G. Stockley [1]*

[1]*Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK*
[2]*York Centre for Complex Systems Analysis, University of York, York YO10 5DD, UK*
[3]*Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxon OX11 0FA, UK*

Received 20 May 2011;
received in revised form
20 July 2011;
accepted 26 July 2011
Available online
3 August 2011

Edited by R. Huber

Keywords:
STNV;
SELEX;
genome;
assembly;
signals

Using a recombinant, *T*=1 Satellite Tobacco Necrosis Virus (STNV)-like particle expressed in *Escherichia coli*, we have established conditions for *in vitro* disassembly and reassembly of the viral capsid. *In vivo* assembly is dependent on the presence of the coat protein (CP) N-terminal region, and *in vitro* assembly requires RNA. Using immobilised CP monomers under reassembly conditions with "free" CP subunits, we have prepared a range of partially assembled CP species for RNA aptamer selection. SELEX directed against the RNA-binding face of the STNV CP resulted in the isolation of several clones, one of which (B3) matches the STNV-1 genome in 16 out of 25 nucleotide positions, including across a statistically significant 10/10 stretch. This 10-base region folds into a stem–loop displaying the motif ACAA and has been shown to bind to STNV CP. Analysis of the other aptamer sequences reveals that the majority can be folded into stem–loops displaying versions of this motif. Using a sequence and secondary structure search motif to analyse the genomic sequence of STNV-1, we identified 30 stem–loops displaying the sequence motif AxxA. The implication is that there are many stem–loops in the genome carrying essential recognition features for binding STNV CP. Secondary structure predictions of the genomic RNA using Mfold showed that only 8 out of 30 of these stem–loops would be formed in the lowest-energy structure. These results are consistent with an assembly mechanism based on kinetically driven folding of the RNA.

Introduction

Single-stranded RNA (ssRNA) viruses exist in every kingdom of life and are major pathogens of humans, animals and plants.[1] There are now many examples of these viruses whose structures have been determined at high resolution by X-ray crystallography or at medium resolution using cryo-electron microscopy.[2–5] In the majority of cases, the principal component of the virion, namely, the viral genomic RNA, is missing from the density

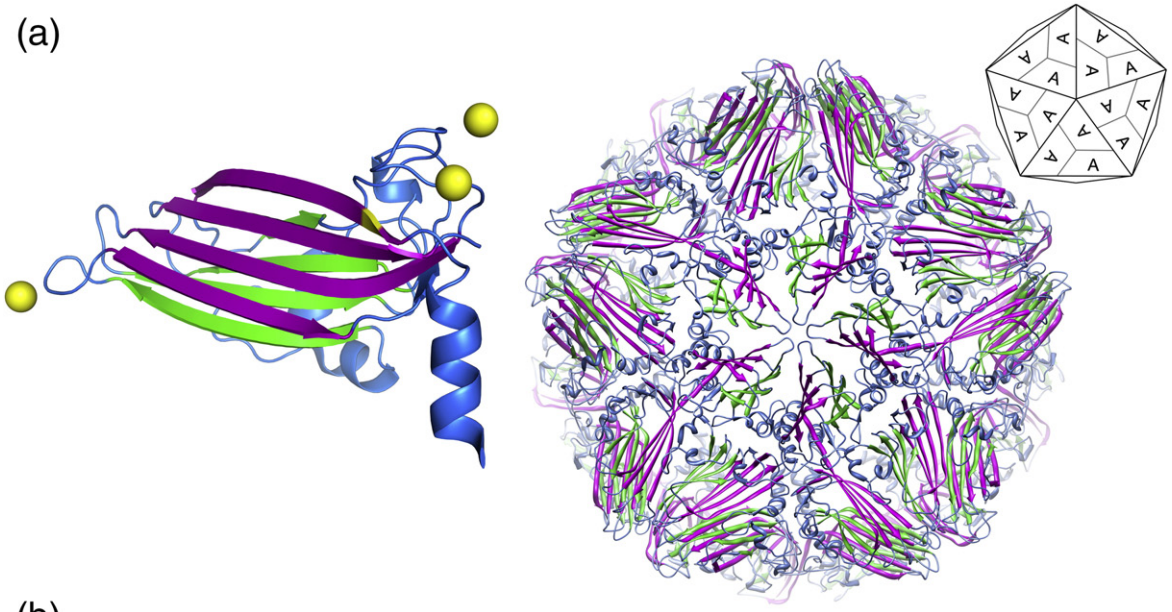
*Corresponding author. E-mail address: p.g.stockley@leeds.ac.uk.

† D.H.J.B. and S.W.L. contributed equally to this work.

Present address: S. W. Lane, Enotria Winecellars Ltd., 4-8 Chandos Park Estate, Chandos Road, London NW10 6NF, UK.

Abbreviations used: ssRNA, single-stranded RNA; CP, coat protein; STNV, Satellite Tobacco Necrosis Virus; VLP, virus-like particle; EDTA, ethylenediaminetetraacetic acid; TEM, transmission electron microscopy.

maps. This has created a protein-centric view of assembly for an important class of pathogens. Recently, we have shown for the bacteriophage MS2 that this viewpoint can be misleading. In that case, the RNA plays multiple functional roles during assembly and possibly disassembly/uncoating that are important for the duration and efficiency of the viral life cycle.[6–11] These observations do not



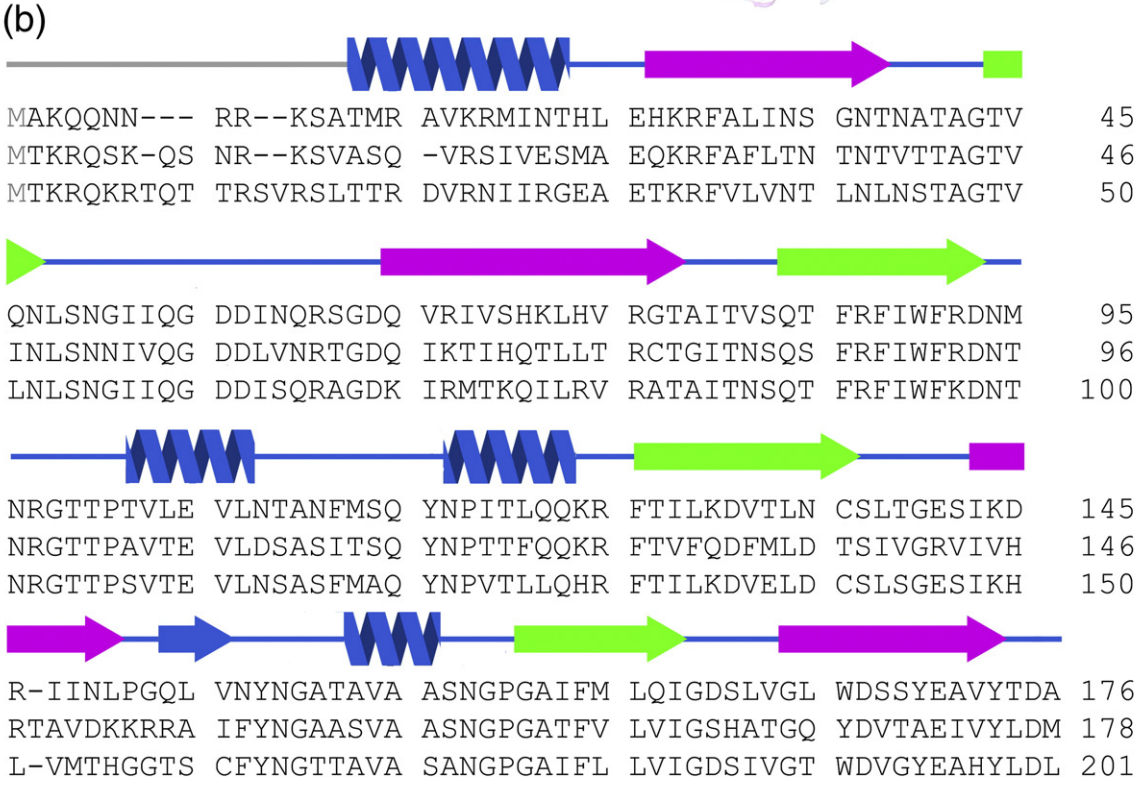


Fig. 1. Architecture of STNV. (a) shows the secondary structural elements of a single subunit of STNV (Protein Data Bank ID 2BUK), with the antiparallel β-sheets shown in magenta (BIDG) and green (CHEF) and the helical regions shown in blue. The three classes of associated calcium ions are shown as yellow spheres. The X-ray structure of the STNV $T=1$ structure is shown alongside with an icosahedron as a guide to its symmetry. (b) shows the amino acid sequences as single-letter codes of the STNV-1 (upper), STNV-2 (middle) and STNV-C (lower) CPs. Shown above the sequences are the positions of the secondary structure elements in STNV-1 from (a).

invalidate the observations that protein capsids can form in the absence of RNA or that heterologous RNAs can be packaged. It does, however, suggest that many such viral RNAs are under selective pressure to aid the formation of the virion. If RNA–coat protein (CP) contacts are important for disassembly/assembly, why are they not visible in electron density maps? Here, again, the example of MS2 is useful. There is frequently little or no electron density for RNA in the high-resolution structures of this and a number of other RNA phages and viruses.[1,12,13] In contrast, in moderate-resolution cryo-electron microscopy structures, substantial fractions of the genome are often seen.[2,4,5] It appears that icosahedral averaging to high resolution often eliminates much of the pseudo-icosahedrally organised, intrinsically asymmetric genome. For MS2, the sequence specificity of CP–RNA interaction has long been known. A high-affinity complex can form between a CP dimer and a 19-nt stem–loop (TR) that serves both as a translational operator of the replicase cistron and as the trigger of assembly of the $T=3$ shell. This has become a paradigm of sequence-specific RNA–protein interaction.[4,6,7,10,14–17]

Recently, we determined the structure of the $T=1$ capsid of Satellite Tobacco Necrosis Virus (STNV) produced in *Escherichia coli* as a recombinant protein from a synthetic mRNA (Fig. 1).[18,19] The structure reported in the adjoining paper shows that this recombinant capsid packages its mRNA, which forms ordered RNA duplexes of ~3 bp that are in contact with the inner surface of the capsid and the N-terminal helices of CP subunits.[20] No obvious sequence specificity has been reported in the interaction between STNV CP and its RNA, raising the question of how wild-type virions package their RNAs.

We have used the recombinant STNV system to show that assembly in *E. coli* is dependent on the N-terminal region of the CP. The capsids can be reversibly disassembled, allowing us to prepare immobilised CPs as selection targets for RNA aptamers.[21,22] The majority of the selected aptamer sequences match to a number of different sites within the STNV genome, suggesting that there is some sequence preference of interaction. The SELEX identified a 4-bp stem topped by a tetra-loop, consistent with the RNA density in the crystal structure.[20] The genome may therefore contain multiple copies of similar sites. The implications for assembly of the $T=1$ capsid are discussed.

Results and Discussion

Disassembly/reassembly of *T*=1 STNV virus-like particles

In order to generate a suitable target for SELEX, we first determined conditions for virus-like particle (VLP) disassembly and reassembly *in vitro*, encouraged by the observation that VLPs assemble spontaneously following expression of the CP in *E. coli*.[20] Disassembly of the purified STNV VLPs was attempted via two methods known to work for other ssRNA viruses. Addition of two volumes of glacial acetic acid leads to disassembly of the RNA bacteriophage MS2,[23] while high salt (1 M) and pH (9.0) buffer in the presence of ethylenediaminetetraacetic acid (EDTA) (3 mM) successfully dissociates other plant viruses, such as Turnip Crinkle Virus.[24] Only the latter led to loss of the $T=1$ VLPs following overnight dialysis, as judged by negatively stained transmission electron microscopy (TEM) (Fig. 2). Dialysis of these disassembled species against the starting buffer (50 mM Hepes–NaOH, pH 7.5, and 3 mM $CaCl_2$) yielded $T=1$ particles by TEM, showing that disassembly and reassembly of the VLPs are possible. Reassembly reactions containing only purified CP but no RNA did not result in the appearance of $T=1$ particles, implying that RNA–protein interaction is essential for assembly under these conditions.

In order to characterise the mechanism of assembly in more detail, we assayed various states of the VLP via sedimentation velocity experiments (Fig. 2 and Table 1). The VLP from *E. coli* has an $S_{20,w}$ of ~42 S, as does the reassembled material, comparable to that of the native STNV-1 virion, $S_{20,w}=38$–45 S.[25] The disassembled material, however, shows heterogeneous peaks ranging from ~1.5 to 30 S, with a sharp peak at ~1.5 S. RNA-free STNV-1 CP, purified as described in Materials and Methods, has a principal component of $S_{20,w}=2.037$ S, with an estimated molecular mass of 18,033 Da, presumably corresponding to a monomer. A similar peak is present in the disassembled VLP sample. The RNA-free STNV-1 CP sample also shows <1% of species with $S_{20,w}$ of ~4 and 6 S, equivalent to trimer and pentamer, respectively, by sedimentation velocity assay. Equilibrium sedimentation of this sample shows that there is no significant self-aggregation under these conditions (data not shown).

As with most ssRNA plant viruses, the STNV CP subunit has an extended, positively charged N-terminal region (Fig. 1). For many of these viruses, this feature plays a vital role in assembly.[24] Residues 1–11 are not visible in the electron density map of the STNV crystal structure;[26,27] thus, they are either disordered or very flexible. The electron density map shows that the ordered region starts with Thr12 (Fig. 1) at the N-terminus of the first helix. We determined the mass of the recombinant subunit using mass spectrometry. The observed mass (21,584.17 ± 0.67 Da) is consistent with that expected for STNV-1 CP minus the N-terminal methionine (21,583.52 Da).

In order to determine whether the N-terminal region in STNV CP plays a role similar to that in the

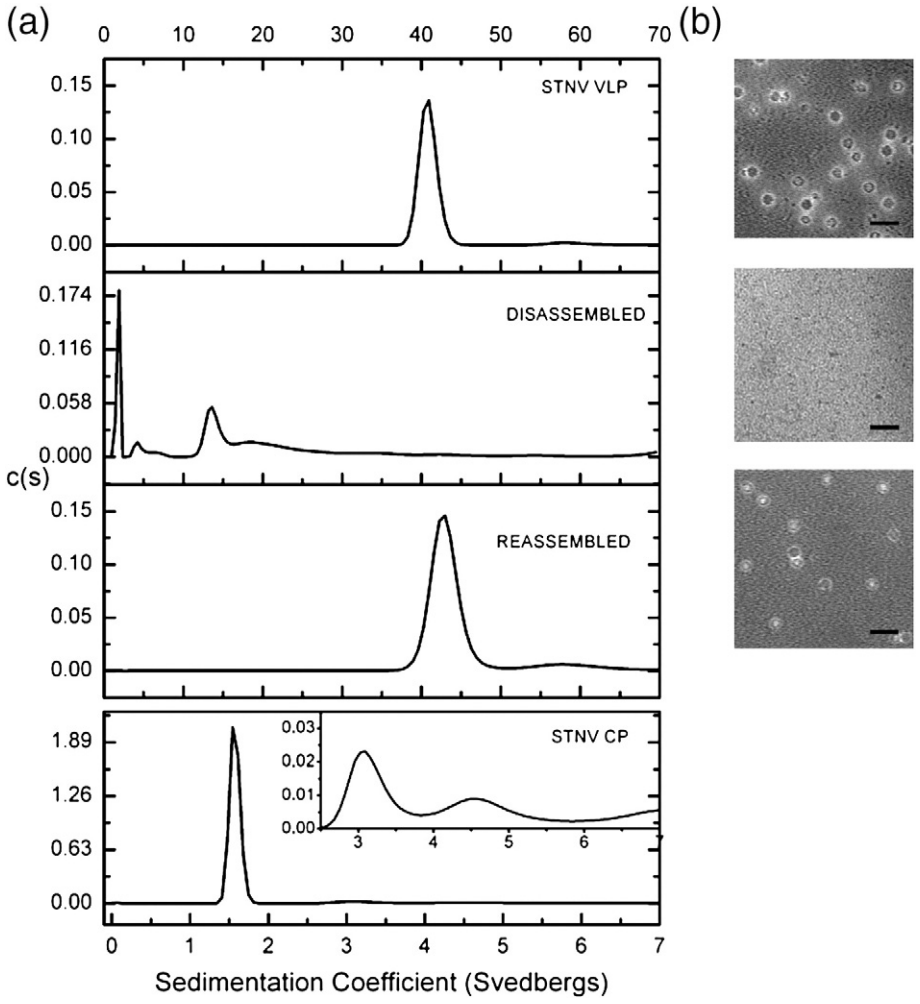


Fig. 2. *In vitro* disassembly and reassembly of the VLP. (a) Normalised $c(s)$ *versus* (S) plots of, from top to bottom, recombinant STNV VLPs, disassembled VLPs, reassembled VLPs and purified STNV-1 CP. The inset in the latter shows a rescaled $c(s)$ axis from 2.5 to 7 S highlighting the minor species present. STNV VLPs were dissociated and reassembled by dialysis and analysed by analytical ultracentrifugation, as described in Materials and Methods. STNV-1 CP was purified under dissociation conditions by ion-exchange chromatography. (b) TEM micrographs of the samples shown on the left; the scale bars represent 50 nm.

Table 1. Calculated hydrodynamic parameters for STNV using analytical ultracentrifugation

Sample	S_{exp}	$S_{20,w}$	f/f_o	Molecular mass (MDa)
Native	40.7	41.7	1.15	1.516 ± 0.076
Reassembled	43.0	43.2	1.14	1.585 ± 0.079416
Purified CP	1.6	2.0	1.3	0.001906 ± 0.0001

S_{exp} is the observed sedimentation coefficient in the following buffers: native: 50 mM Hepes–NaOH (pH 7.5) and 200 mM NaCl at 20.0 °C; reassembled: 50 mM Hepes–NaOH (pH 7.5) and 3 mM $CaCl_2$ at 20.0 °C; and purified CP: 50 mM Hepes–NaOH (pH 9) and 3 mM EDTA at 10.0 °C. $S_{20,w}$ is the sedimentation coefficient corrected for buffer conditions, as described in Materials and Methods. f/f_o is the frictional coefficient. A spherical particle has an f/f_0 value of 1.

CPs of larger plant viruses, we created a truncated protein (Δ15) lacking residues 1–15 (see Materials and Methods). Ala15 is one turn from the start of the N-terminal helix (Fig. 1). SDS-PAGE of extracts from cells expressing the variant showed that the yield of the Δ15 protein was similar to that of wild-type. However, when the purification protocol used for the wild-type VLP was applied to this mutant, the final size-exclusion column profile showed two peaks. The smaller of these corresponds to free Δ15 monomers, and the larger, to aggregates of protein contaminants. It appears, therefore, that deletion of the N-terminal region prevents association with RNA and formation of $T=1$ particles. In some reassembly experiments, the CP sample contained a fraction of proteolysed material, presumably lacking the N-termini. SDS-PAGE of the

reassembled $T=1$ products with this sample showed no evidence of truncated subunits, implying that only intact subunits form VLPs and confirming the importance of the N-terminal region for interaction with RNA.

Identification of preferred RNA sequences for STNV-1 CP

In the adjoining paper,[20] Lane *et al*. have shown that the recombinant STNV VLP expressed from a synthetic mRNA in *E. coli* forms $T=1$ capsids containing RNA, a significant fraction of which is that mRNA, implying packaging selectivity *in vivo* with respect to cellular RNAs. The crystal structure also shows that there must be multiple RNA–protein contacts in the VLP to account for the ordered electron density corresponding to segments of bound double-stranded RNA. In order to determine how these VLPs package RNA, we used SELEX to determine whether there is a preferred RNA sequence involved in these interactions. *In planta* and for many animal ssRNA viruses, viral genome packaging is often viewed as occurring on nascent RNA strands, as they are released from the RNA-dependent RNA polymerase, independent of sequence-specific RNA–protein contacts. The precise details of STNV CP–mRNA complex formation in *E. coli* are necessarily different from those in the plant cell with full-length genomic RNA. For the VLPs used here, for instance, the synthetic mRNA is produced by T7 RNA polymerase,[28] not by the Tobacco Necrosis Virus RNA-dependent RNA polymerase.

For SELEX experiments, it is important to define the selection target carefully. For MS2, an appropriate selection target is the CP dimer, since this is the clear structural building block within the $T=3$ capsid.[10,29–32] It is not obvious, however, what the basic capsomere of the STNV capsid should be. We therefore adopted a generic strategy that should, in principle, work for any size of capsomere (Fig. 3). STNV VLPs were partially biotinylated using EZ-Link® Sulfo-NHS-LC-LC-Biotin (Pierce Biotechnologies), which modifies solvent-exposed lysine residues lying principally on the outer surface of the

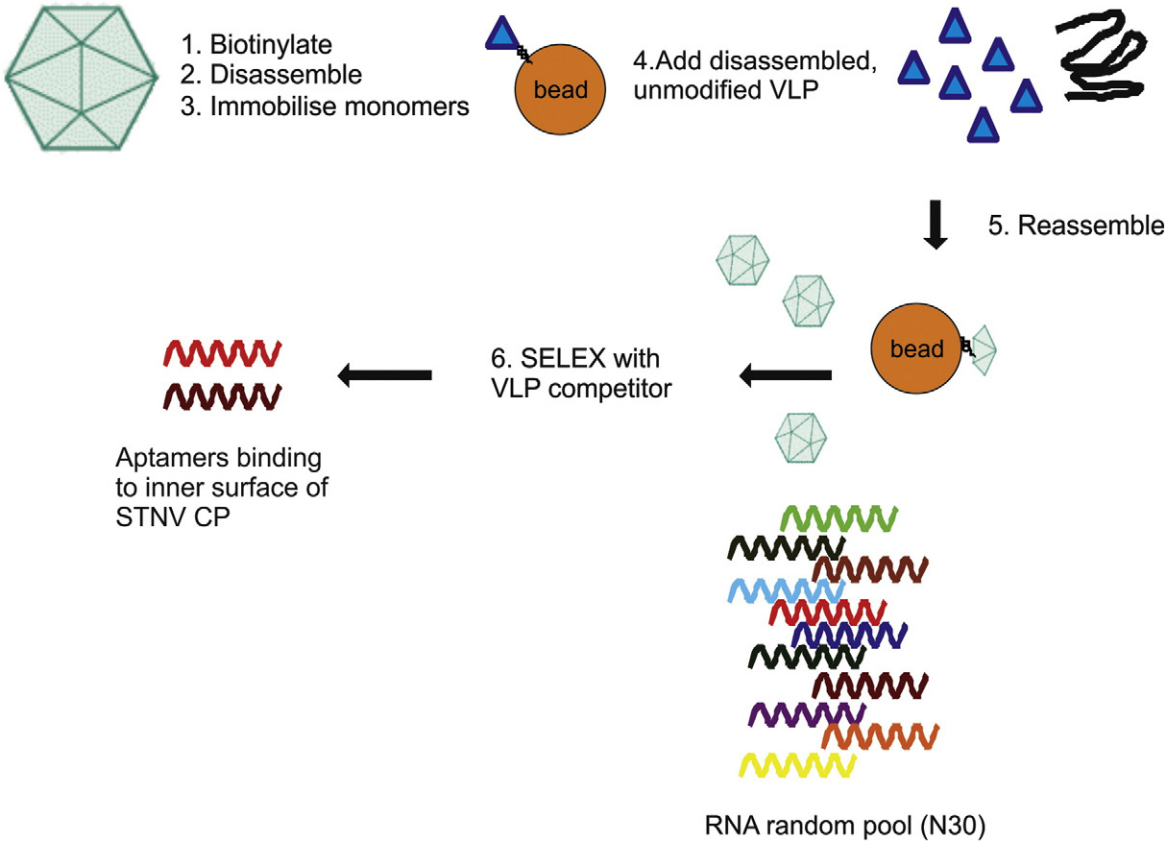


Fig. 3. Scheme for SELEX. VLPs were initially modified with EZ-Link® Sulfo-NHS-LC-LC-biotin before being disassembled to monomers. These were then immobilised on streptavidin-coated beads before addition of unmodified, dissociated VLPs, and then the mixture was transferred to reassembly buffer, creating a range of partially assembled, immobilised SELEX targets. SELEX rounds included intact, recombinant VLPs "free" in solution to prevent recovery of aptamers directed at epitopes on the outer surface of the immobilised capsomeres.

Table 2. Sequences of 10th round STNV-1 CP aptamers

A1	CGTCATGAATAGAGTTTTTTGCGGTCAGCA
A2	GTGTTTATCTTGTCACTAGGTACGGGTATG
A5	CGCGACCATTGAAATCACTTCTGGGGTTAA
B1	GTGTAATTTGGTTTATTCTTTCAATGGTGA
B2	CATTACAACTATCACTTCTCCTC
B3	CCTTTTCAAGACATGCAACAATGCACACAG
B4	CGTTTTAGTTTTCGGGTATTAATAGGCGTG
B5(&F2)	TATGAGTTGATTTGTGGTTTTTGGTAACGT
C1	CTTGATTAACTAAAGGCTAGCAAACAGCG
C2	AACATTTGGATGTATGCCTTAGTTACCTAG
C3	CACAGGTAAGAATCCAGCCCTATACTAACA
C4	TGGACACCCTAAAAATATACGGCTCCCTGG
C5	ACATGCCTTTATTTTGTCTTGGATCTTGAC
D1	GTCTCCATGAACAATATTGTAGCTACAGGTC
D5	CTGTGTAGTTTGAAATCATTATCGCCAACA
E5	GAGCAATCGGGATACCTTCACATGATGTGC

subunits of the VLP, for example, Lys27. These modified $T=1$ shells were then disassembled as described above and incubated with MyOne™ streptavidin-coated Dynabeads® (Invitrogen), allowing monomeric CP subunits to be immobilised. The beads were washed and then added to a sub-stoichiometric solution of disassembled, unmodified VLPs in reassembly buffer. This reaction therefore includes the RNA that was packaged in *E. coli*. The rationale behind this protocol is that the presence of RNA should allow at least partial capsid assembly to occur, including on the immobilised subunits. The ratio of biotinylated to unmodified CP was controlled to ensure that there was insufficient CP to allow all immobilised subunits to form complete capsids. Partially formed capsids on beads should therefore display exposed RNA binding sites on CP surfaces that would be hidden in $T=1$ shells. Although, in principle, some of these RNA binding sites will be occupied by the RNA encapsidated in *E. coli*, some sites are likely to be free. In addition, the SELEX procedure is such that tighter-binding RNA sequences would be likely to displace weaker sequences during selection, especially as only aptamer sequences are amplified and recovered at each round. These partially assembled, immobilised targets were therefore used in 10 rounds of SELEX in assembly buffer. In order to remove aptamers directed at the outer surface of the particle, we added unmodified $T=1$ VLPs to the selection step.

The resulting 10th round aptamer pool was cloned, and 17 individuals were sequenced (Table 2), 16 of which were unique sequences. When their sequences were compared using the "AliBee" multiple sequence alignment algorithm‡, a number of sequence motifs could be identified, but there was no clear consensus. This differs from the MS2 SELEX experiment where a clear consensus emerged.[32] However, when these sequences are compared to the genomic sequence of STNV-1, several statistically significant matches to different regions of the genomic RNA are observed (Supplementary Figs. 1 and 2). Of these, one aptamer (B3) has 16 nt that match the STNV-1 genome from nucleotide positions 57–81, that is, 16/25, including a contiguous 10-nt stretch (Fig. 4a). This degree of identity is unlikely to occur by chance, implying that this is a preferred STNV CP binding site. Mfold analysis of B3, as well as the genomic region to which it matches, suggests that these sequences can fold into stem–loops (Fig. 4b), which are common protein recognition motifs. The nucleotides immediately adjacent to the region of sequence identity can form an additional Watson–Crick base pair in both B3 and STNV-1, extending the length of the stem to 4 bp, hence increasing the significance of the match. The electron density map of the recombinant VLP is consistent with an RNA binding site containing at least a 3-bp stem, and the region above the stem and below the capsid inner surface has a volume appropriate to accommodate at least a 4-nt loop.[20] A similar but significantly less stable stem–loop could be formed on the mRNA generated from the recombinant gene.

In order to explore whether the B3 motif was functional, we determined its secondary structure using enzymatic structure probing in the presence and absence of CP (Fig. 5). The results are consistent with the predicted secondary structure and show that the B3 stem–loop is protected from enzymatic digestion in the presence of CP, while other stem–loops in the flanking sequences are not. RNA fragments encompassing the selected B3 aptamer (30-mers) also trigger assembly of CP into $T=1$ VLPs *in vitro* using the reassembly protocol described above (data not shown).

As a test of the functional significance of this interaction *in vivo*, we carried out site-directed mutagenesis of the synthetic gene to alter the B3-like mRNA stem–loop to disrupt (SL–) or mimic (SL+) the genomic version of the stem–loop, that is, altering the base-pairing potential without altering the encoded CP (Fig. 4c). These mutants were used to produce VLPs in *E. coli*. In each case, the yields of VLP and the $A_{260/280}$ values of the purified particles were similar to those of the original mRNA, suggesting that they had packaged equivalent amounts of RNA. This suggests that, unlike the TR site in MS2,[10] this B3-like stem–loop is not a unique assembly initiation site.

Another possible explanation of the role of the RNA–CP interaction is that multiple stem–loops contribute to both assembly and particle stability. This could account for the fact that multiple matches between the aptamers and various sites on the genome are observed for STNV-1 (Supplementary Fig. 2). In this type of assembly model, multiple stem–loops having sequence similarity rather than strict identity would be important, and assembly

‡http://www.genebee.msu.su/services/malign_reduced.html

would therefore not be compromised in the mutagenesis experiment described above. The next best alignment is for aptamer D1, which matches at two sites within the STNV-1 genome with 17/23 or 18/23 identities, both with a longest continuous match of 5/5. Mfold predicts that D1 can also form a stem–loop, displaying the same ACAA motif as B3, although in the context of a five-base loop (AACAA). The other aptamers (Table 2) were analysed for their ability to form stem–loops. Four of them (A2, B3, B4 and C3) are predicted to form stem–loops with 4-nt AxxA motifs, and a further five (A5, C1, D1, D5 and E5) display this motif in the context of a 5-nt loop (Fig. 4d). Thus, 11 out of 16 aptamers display the AxxA in a loop of 4 or 5 nt at the end of a base-paired stem. We postulated that the AxxA motif might be important for CP binding and therefore asked whether the remaining six aptamers can form stem–loops encompassing an AxxA motif if the number of nucleotides allowed in the loop is increased to six. This identified C4 as able to form such a stem–loop. We then further extended the analysis to include the fixed flanking regions that are part of the RNA transcripts used for selection, that is, not just the selected regions listed in Table 2. This identified a further three aptamers that display AxxA in either a 4-nt loop (B5) or 6-nt loops (C2, A1 and C5). Only aptamers B1 and B2 do not conform to this pattern, and one of these is much shorter than 30 nt in length.

From this analysis, we constructed a set of six stem–loop motifs, with between four and six degenerate nucleotides in the loop and an allowance for non-Watson–Crick base pairing in the stem (Fig. 6), and used these to probe the STNV-1 genome for matches. This revealed 30 potential stem–loops (SL1–SL30, Fig. 6), enough to account for the occupancy of RNA in the sites observed in the crystallographic maps.

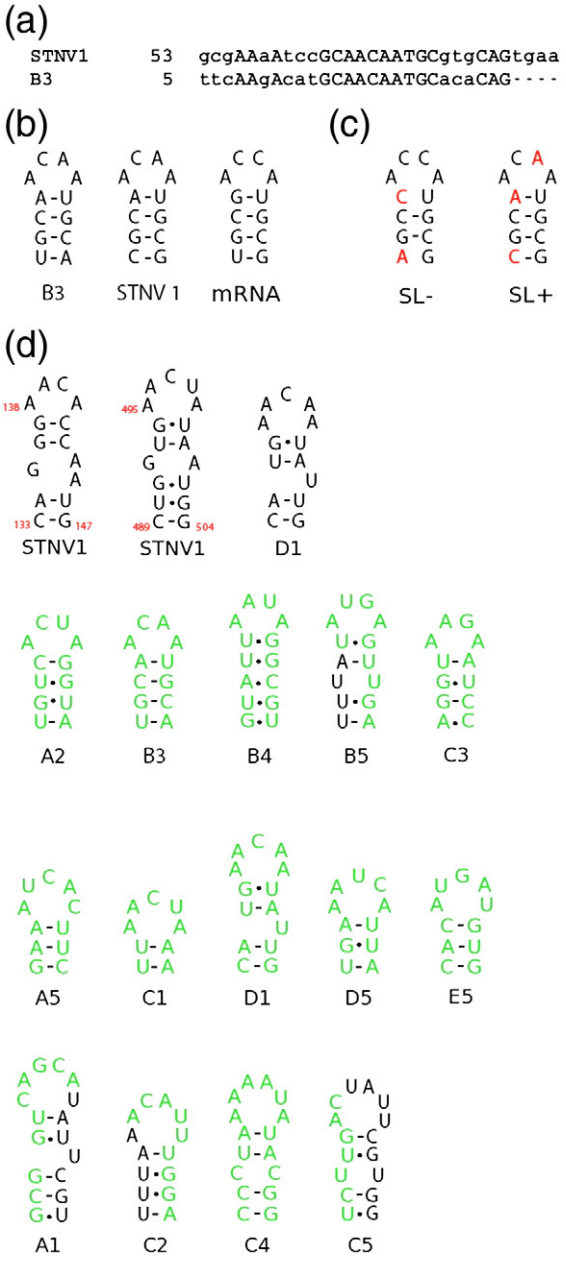


Fig. 4. Identification of a common. stem–loop in aptamer B3, STNV-1 and mRNA. (a) Best match of aptamer B3 (Table 2) with the STNV-1 genome determined using the AliBee multiple sequence alignment tool. Capital letters indicate identities between the two sequences. (b) Stem–loop motifs obtained using the Mfold program, from the contiguous 10-nt match in (a), plus the flanking bases in B3 that form an additional base pair at the bottom of the stem. Also shown are the matching stem–loops that can form in the STNV-1 genome and the mRNA. (c) Mutations made to the B3 stem–loop in the mRNA in order to disrupt (SL−) or improve (SL+) its match to the genomic stem–loop. Mutated residues are shown in red. (d) The top line shows the secondary structure of the next best aptamer match (D1) to the STNV-1 genome after B3, which matches to two sites within the genome, all of which can be folded into stem–loops displaying the AXXA motif. The lower half of the figure shows Mfold secondary structures of the 14 aptamers able to form stem–loops displaying the AXXA motif in four-, five- or six-base loops. Regions of selected sequence are shown in green, and fixed flanking regions are in black.

The roles of STNV stem–loop/CP interactions: Implications for assembly

If STNV CP has some preference for binding the potential stem–loops identified above, its assembly mechanism must be a hybrid of the two extremes mentioned previously. Since there appears to be no single high-affinity assembly initiation site, a series of stem–loops capable of CP recognition could facilitate packaging as the newly replicated genome leaves the polymerase. This is a plausible mechanism for STNV assembly. Many other RNAs by chance

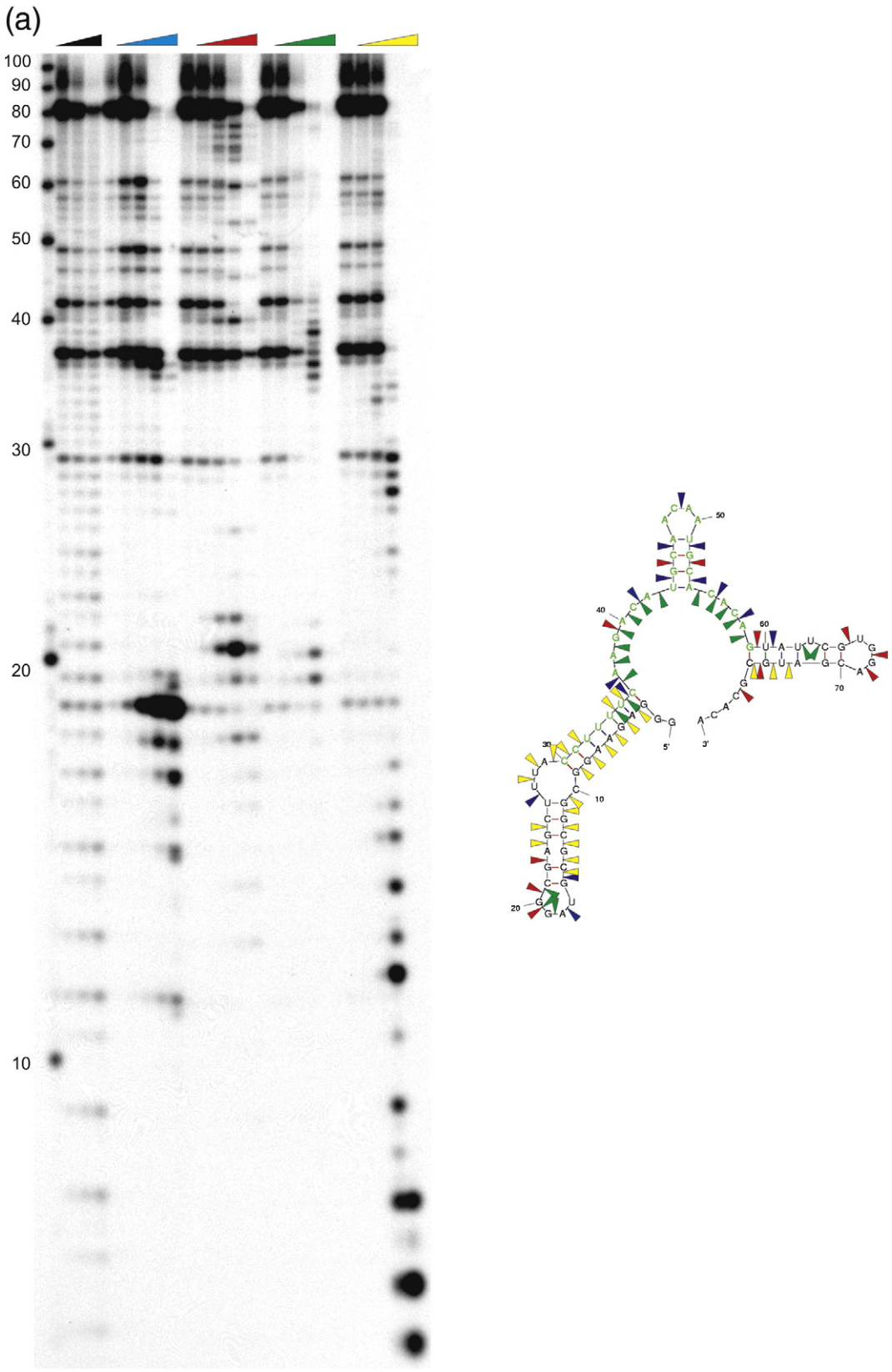


Fig. 5 (*legend on next page*)

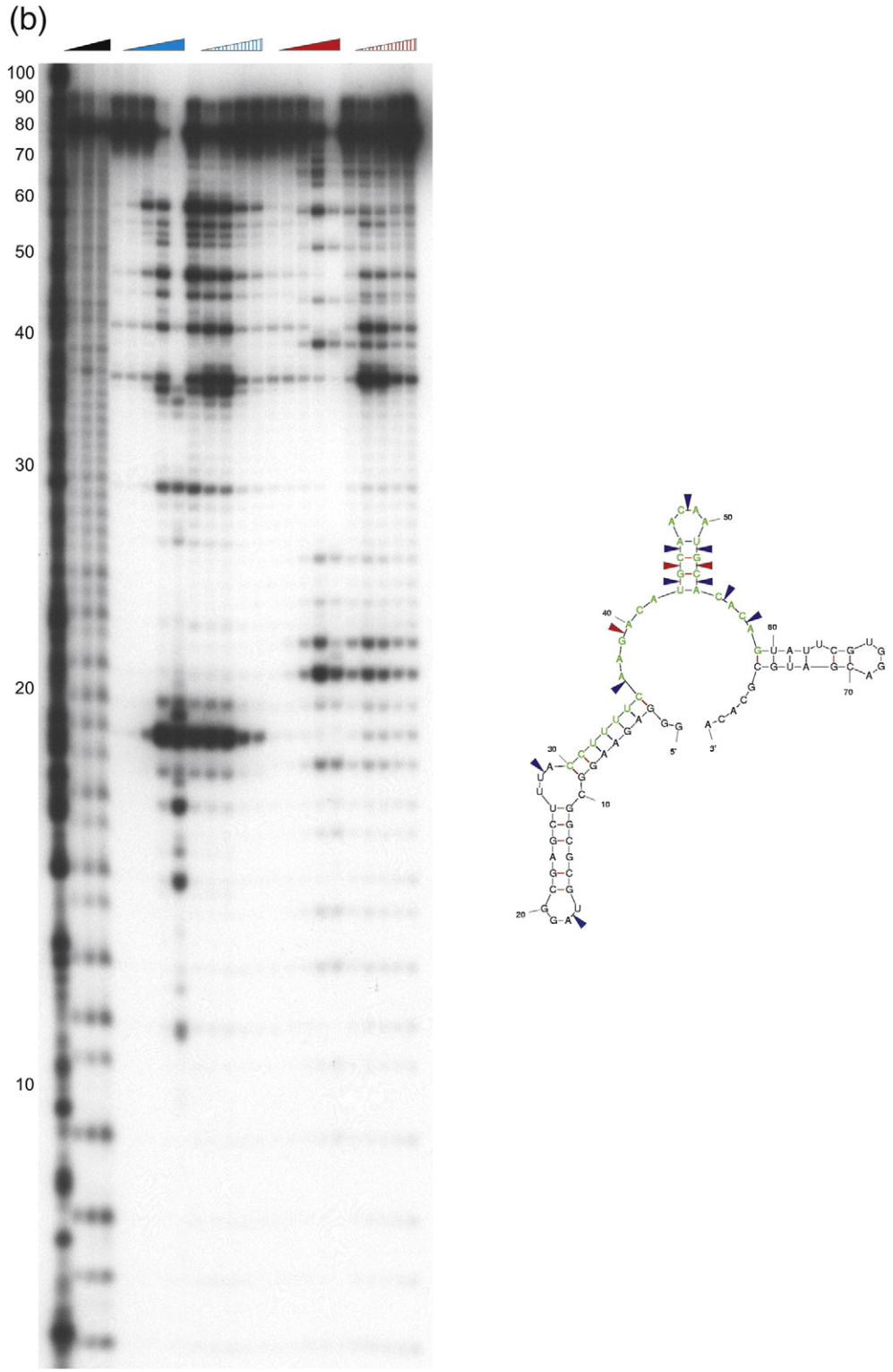


Fig. 5. B3 aptamer is bound by STNV-1 CP. (a) The secondary structure of aptamer B3 in solution, predicted using Mfold (right), was consistent with the results of denaturing polyacrylamide gel electrophoresis (left) after enzymatic digestion with RNases. Cleavage sites are shown mapped onto the predicted secondary structure [RNase A (shown in blue), RNase T1 (red), RNase V1 (yellow) and S1 nuclease (green)]. Bases in the initially randomised region are shown in green. (b) The RNase A and T1 digestions were repeated in the presence of a 2- to 16-fold molar excess of the STNV-1 CP. Sites that were protected from cleavage are indicated.

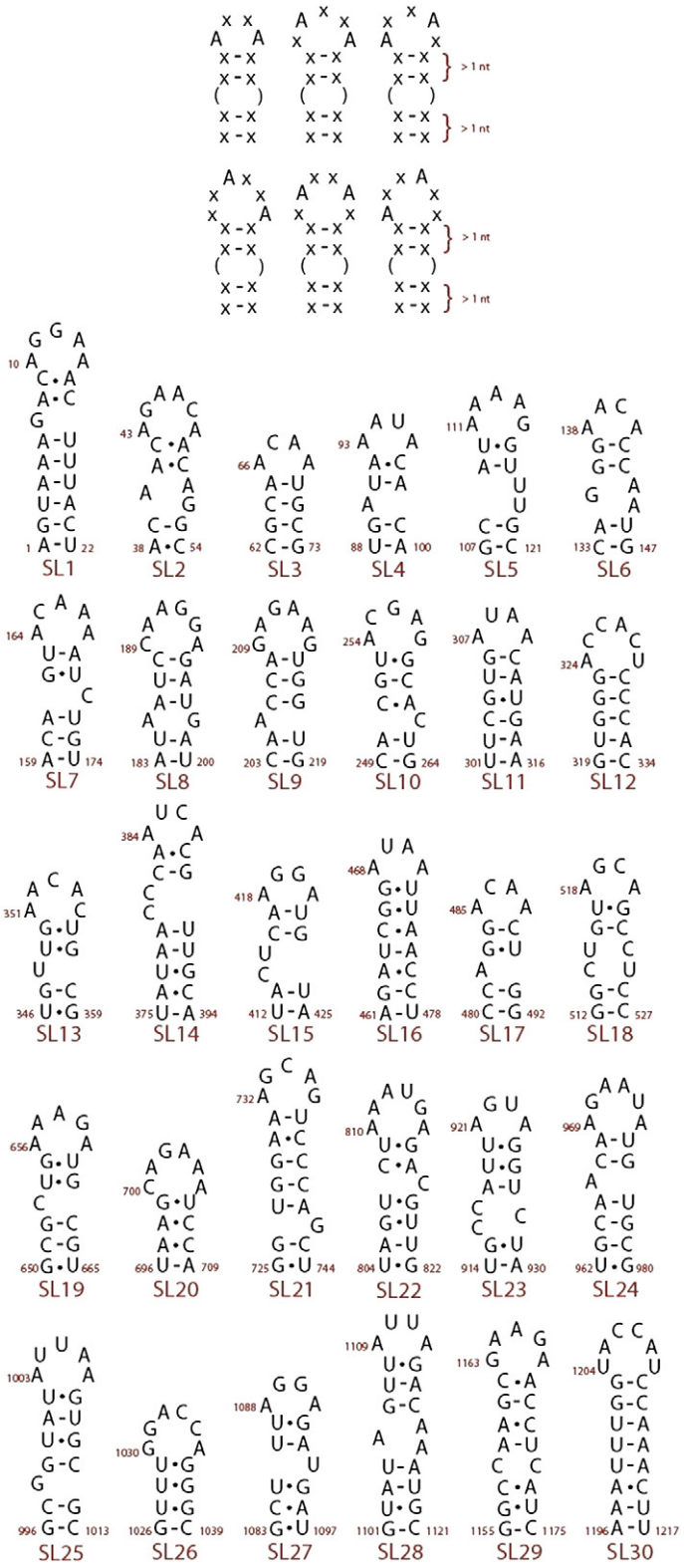


Fig. 6. Stem–loops in STNV-1 with motifs implied by aptamers. The search motifs used to locate stem–loops within the STNV-1 genome. These are based on the selected aptamers (Table 2) and subsequent alignments to the STNV-1 genome. The 30 stem–loops identified within the STNV-1 genome (SL1–SL30) are shown below. The red numbers show their positions within the viral RNA.

will display one of the preferred RNA sequence motifs, accounting for nonspecific RNA packaging. Both the STNV-2 and the STNV-C genomes, as well as the mRNA, contain multiple copies of stem–loops displaying the AXXA motif, suggesting a common packaging mechanism (Supplementary Fig. 3). These RNAs have an advantage over random RNAs since they contain multiple copies of this recognition

motif. Their encapsidation would thus benefit from the cooperativity of protein–protein interactions in the assembling capsid of STNV-1.

Testing this hypothesis directly by mutagenesis of all the SL stem–loops would be a complex task. We can, however, investigate the differences between the secondary structures of STNV-1 genomic RNA predicted under either thermodynamic or kinetic constraints (Fig. 7). The kinetically folded structures were constrained to contain all putative stem–loops, SL1–SL30, and to restrict potential base pairing to local regions of no longer than 50 nt. The thermodynamic structure, obtained without imposing these constraints, is only predicted to contain eight members of this set. Mfold predicts a free-energy difference of ~190 kcal/mol in favour of the thermodynamically folded structure. However, if all SL1–SL30 stem–loops make favourable contacts with CPs of binding free energy of ~8 kcal/mol, then the kinetic model would be favoured even before the effects of protein–protein interactions between CPs in the capsid are considered. This compares with a ΔG of ~10 kcal/mol for the interaction of the MS2 CP with TR.[10,33] Since the aptamer sequence matches would be unlikely to occur in a genome folding thermodynamically, these results strongly favour the kinetic folding model.

In principle, the proposed assembly model could be confirmed by determining the secondary structure of the genomic RNAs *in situ* in the capsid. There are ways to do this, and these experiments, although complex, are underway. Having a complete secondary structure model would also allow a complete model of the STNV-1 virion to be built, as has been attempted for Satellite Tobacco Mosaic Virus.[34–36]

Materials and Methods

Purification and disassembly/reassembly of STNV VLPs

Recombinant STNV $T=1$ VLPs were expressed from pET22b-STNV and purified using a slight variation of the initial protocol.[20] Purified VLPs (~7 mg ml^{-1}) were disassembled by dialysis against disassembly buffer [50 mM Hepes–NaOH (pH 9), 1 M NaCl and 3 mM EDTA] at 4 °C overnight with gentle stirring. Reassembly of the monomeric CP (~1 mg ml^{-1}) was achieved by dialysis against reassembly buffer [50 mM Hepes–NaOH (pH 7.5), 100 mM NaCl, 3 mM $MgCl_2$ and 3 mM $CaCl_2$] at 4 °C overnight with gentle stirring. After dialysis, up to 90% of the material sedimented with approximately the same S value as the starting material (Table 1).

SELEX target preparation

For isolation of RNAs that recognise and bind specifically to the capsid interior, the selection target ideally would represent the viral capsomere. We employed a strategy to create a range of assembly intermediates on the way to the $T=1$ capsid, in the hope that the viral capsomere would be sufficiently represented to isolate the desired aptamers. Intact STNV VLPs (500 μl, ~1 mg ml^{-1}) were partially biotinylated using EZ-Link® Sulfo-NHS-LC-LC-biotin (Pierce Biotechnologies), at a final concentration of 0.2 mg ml^{-1}. After incubation (room temperature, 10 min), the reaction was quenched by the addition of 1 ml of 1 M Tris–HCl, pH 8. The partially biotinylated VLPs were then dialysed overnight against disassembly buffer to remove both unreacted biotinylation reagents and to disassemble the VLPs. Biotinylated CP monomers (~0.5 mg ml^{-1}) were immobilised on 1 μm MyOne™ streptavidin-coated Dynabeads® (Invitrogen), according to the manufacturer's protocol. Immobilisation was confirmed by absorbance measurements pre- and post-bead binding. Following extensive washing (in reassembly buffer), derivatised beads were incubated with non-biotinylated STNV monomers at 0.1 mg ml^{-1} for 6 h in reassembly buffer. It was assumed that unmodified monomers would assemble with the bead-bound STNV monomers to provide a range of immobilised assembly intermediates. Again, bead binding (and therefore assembly) was confirmed by measuring the optical density spectrum pre- and post-incubation. Unbound material was removed by extensive washing in assembly buffer. At this ratio of modified to unmodified CP, complete reassembly on the immobilised monomers would be impossible.

In vitro selection—SELEX

The initial double-stranded DNA SELEX library consisted of ~10^{15} sequences, each containing an N30 random region flanked by fixed primer regions, one of which carried the T7 RNA polymerase promoter. All RNAs were prepared in 50 μl reactions containing a final concentration of 40 mM Hepes–NaOH (pH 7.6), 26 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 2.5 mM ATP, 2.5 mM GTP, 2.5 mM UTP, 2.5 mM CTP, 5 μl double-stranded DNA template, 100 units of T7 RNA polymerase (Roche) and 0.5 units of Yeast Inorganic Pyrophosphatase (Sigma). This mixture was thoroughly mixed and then incubated at 37 °C for 3 h. Template DNA was subsequently removed by addition of 6 μl DNase I buffer [to a final concentration of 40 mM Tris–HCl (pH 8), 10 mM $MgSO_4$ and 1 mM $CaCl_2$] and 4 units of RQ1 DNase (Promega) and incubation at 37 °C for 20 min. The RNA was purified using RNAClean™ resin following the manufacturer's protocol (Agencourt).

The purified RNA pool was added to 300 μg of beads carrying the immobilised, partially assembled STNV VLPs in assembly buffer. Intact, unmodified STNV VLPs (0.1 mg ml^{-1}) were also added to serve as a counter-selection to remove any aptamers directed against the outer surface of the STNV VLP. After incubation at 20 °C for 20 min, the beads were separated and gently washed six times to remove any unbound or intact VLP-bound RNA. Aptamers were then eluted by heating the beads to 95 °C for 10 min in 30 μl RNase-free water. Eluted RNAs were reverse transcribed and PCR amplified as described below.

RT-PCR mix (20 μl) was added to the eluted RNA to give a final concentration of 50 mM Tris–HCl

(pH 8.5), 30 mM KCl, 8 mM $MgCl_2$, 2 μM reverse primer (5′-TGTGCGCATCGTCCACGAATA-3′), 1 mM each dNTP and 20 units of Transcriptor™ reverse transcriptase (Roche). This was incubated at 52 °C for 2 h before adding the PCR mix. PCR amplification was carried out in 70 μl reactions, containing a final concentration of 20 mM Tris–HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM dNTP mix, 500 nM forward primer (5′-TAATTCTAATACGACTCACTA-TAGGGAGAAGGCGGCGCGTAGGCGAGCTTTA-3′), 0.5 μM reverse primer and 2.5 units of Taq DNA polymerase (Invitrogen). Ten cycles of PCR amplification were carried out to limit the emergence of PCR artefacts. This PCR product was used as the template in the subsequent round of selection. A Biomek 2000 automated workstation (Beckman Coulter) was used to carry out 10 rounds of *in vitro* selection as described previously.[21,22] The 10th round pool was cloned into the pGEM-T Easy sequencing vector following the manufacturer's protocols (Promega). A number of aptamer-containing colonies were identified using blue-white screening, the plasmids were isolated and individual aptamers were sequenced (Table 2).

Enzymatic secondary structure probing

Individual aptamer clones were selected and transcribed as described above. Purified RNA (5 μg) was dephosphorylated and 5′ end-labelled with [γ-^{32}P]ATP following standard protocols. This was then phenol/chloroform extracted, ethanol precipitated and resuspended in 50 μl RNase-free water. Samples (1 μl) were digested with various concentrations of RNase A, T1, V1 (Ambion) or S1 nuclease (Invitrogen) following the manufacturers' protocols. Digestion products were run

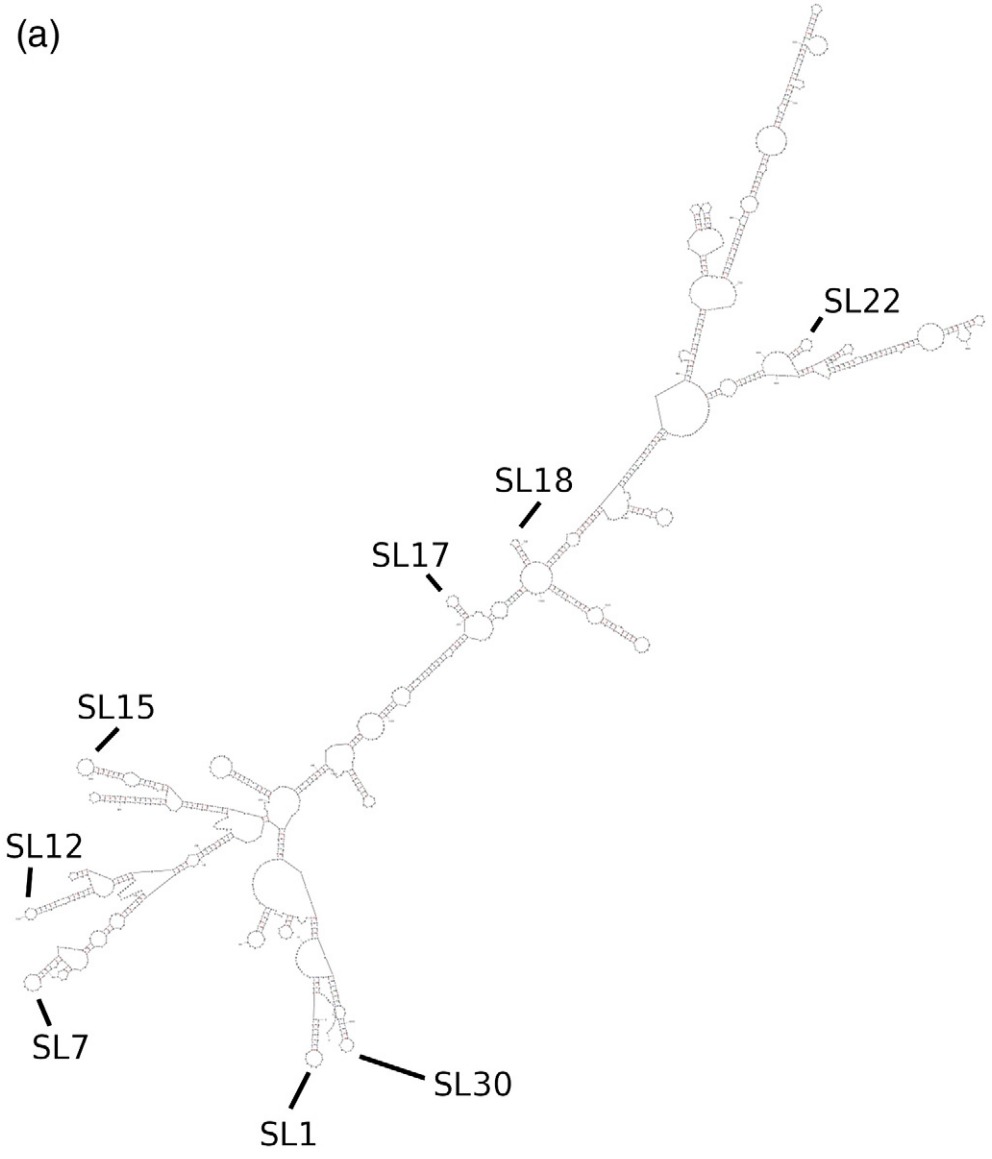


Fig. 7 (*legend on next page*)

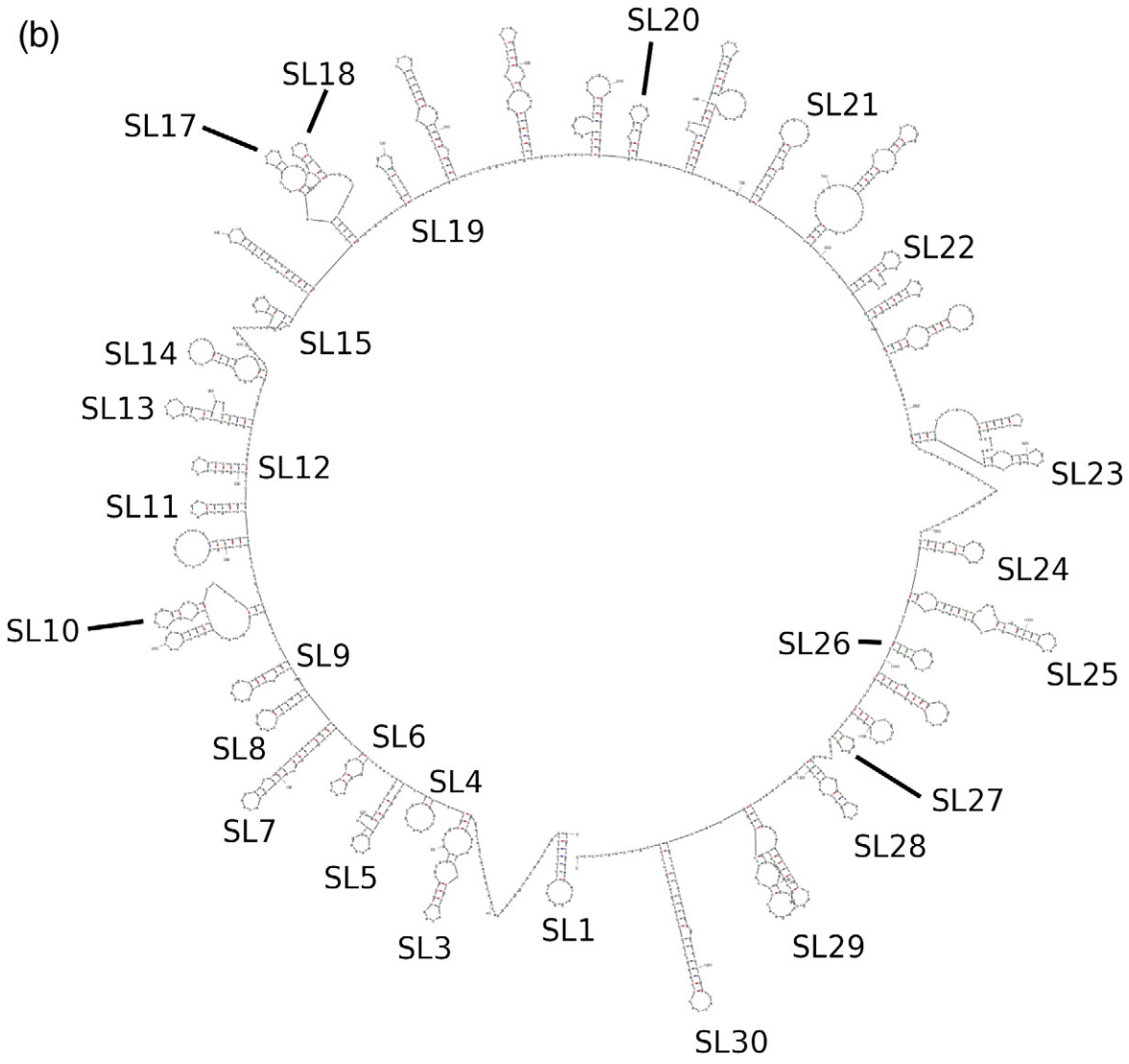


Fig. 7. Predicted secondary structures of STNV-1 RNA. (a) Secondary structure of STNV-1 RNA predicted by Mfold using the default settings, that is, a thermodynamically controlled fold. Stem–loops corresponding to those in Fig. 6 are labelled. (b) Secondary structure of STNV-1 RNA predicted by Mfold using SL1–SL30 as constraints, that is, a kinetically controlled fold. In order to ensure preference for short-range interactions, we limited the program to look for base pairing within a 50-nt window.

on a prewarmed 20% (w/v) denaturing SequaGel™ (National Diagnostics). A radiolabelled Decade Marker™ (Ambion) and an alkaline hydrolysis ladder were also run. Digestion products were imaged by autoradiography.

Nuclease protection assays

Individual aptamers were labelled and digested with RNase A or T1, as described above. Samples were also digested in the presence of a 2- to 16-fold molar excess of unmodified, partially reassembled STNV monomer. As previously, digestion products were run on a 20% (wt/vol) SequaGel™ and visualised by autoradiography.

STNV mutagenesis

In order to produce an N-terminally deleted protein, we designed primers to create a 15-amino-acid deletion from the N-terminus of the protein (Fig. 1). The primers (Δ15 forward primer 5′-GGAATTCCATATGGTGAAACG-CATGATC-3′ and Δ15 reverse primer 5′-CCCAAGCT-TACGCGTCAGTGTAC-3′) contained restriction sites for NdeI and HindIII to allow insertion of the PCR product into pET22b-STNV using standard methods. Single colony transformants were picked and grown in 5 ml of LB, 100 μg ml^{-1} ampicillin and the expression plasmid extracted using the Qiagen miniprep kit. HindIII and NdeI restriction digests were carried out to ensure that the

gene fragment was present before the plasmids were sequenced, which confirmed the truncated CP gene.

Similarly, mutagenesis to weaken (SL−) or strengthen (SL+) the base pairing of the STNV-1 genome region matching aptamer B3 was carried out using the following primers, and the isolation of the correct sequences was confirmed by sequencing. Primers were:

SL− forward, 5′-CGCAAATCAGCCACCATGCGC-3′;
SL− reverse, 5′-GCGCATGGTGGCTGATTTGCG-3′;
SL+ forward, 5′-CGCAAATCCGCAACAATGCGC-GCG-3′; and
SL+ reverse, 5′-CGCGCGCATTGTTGCGGATTTGCG-3′.

RNA structure prediction

RNA secondary structure prediction was performed *in silico* using the program Mfold§.[37] Structure analysis was carried out using aptamer sequences, and matching portions of STNV-1 and synthetic STNV and sections of STNV-1 were run using the program default parameters.

Stem–loop motif search algorithm

The stem–loops in the genomes of STNV-1, STNV-2, STNV-C and the mRNA that match the search pattern shown in Fig. 6a were determined via the following algorithm. A window of 4, 5 or 6 nt in size was translated in increments of 1 nt along each of the sequences. At each step, the algorithm tested whether the two nucleotides at the 5′ and 3′ ends, respectively, of the window were able to base pair. If this was the case, it continued base pairing as far as possible, permitting at most one bubble (of up to 2 nt on either side) in between. If alternative folds were possible, they were all recorded, and their energies were ranked according to the following scoring rule: every G–C base pair incurred a score of −3, A–U incurred a score of −2, G–U incurred a score of −1 and A–C incurred a score of 0. The algorithm then filtered this list (which encompassed about 6000 stem–loops with a 4-, 5- or 6- nt loop in total) for those stem–loops that contained the motif AxxA in the loop portion. If overlapping stem–loops were identified in this way, the stem–loops of lowest energy (according to the scoring rule above) were selected.

Analytical ultracentrifugation

Sedimentation velocity

Samples (0.32 ml) were centrifuged in 1.2-cm-path-length 2-sector meniscus-matching epon centrepiece cell built with sapphire windows in an 8-place An50Ti analytical rotor in an Optima XL-I analytical ultracentrifuge (Beckman Instruments, Inc., Palo Alto, CA 94304). Changes in solute concentration were detected by interference and absorbance scans at 260 and 280 nm. STNV CP was run at a rotor speed of 50,000 rpm at a temperature of 10.0 °C; all other STNV experiments were run at 15,000 rpm and 20.0 °C. Buffer densities and viscosities were calculated using the program Sednterp[38] version 1.09 from Dr. Thomas Laue, University of New Hampshire, USA. Radial absorbance plots were used for fitting to sedimentation profiles with the program Sedfit version 12.1b (2010) using a continuous distribution $c(s)$ Lamm equation model.[39]

Sedimentation equilibrium

Samples (0.125 ml) were centrifuged in 1. 2-cm-path-length 6-sector epon centrepiece cells with sapphire windows in a 4-place An-60 Ti analytical rotor running in an Optima XL-I analytical ultracentrifuge (Beckman Instruments, Inc.) at a temperature of 5.0 °C. Radial absorbance scans (20 replicates; radial step size, 0.001 cm) and interference scans were collected. Scans were judged to be at equilibrium by plotting radial offset *versus* time using Sedfit. The results were analysed using the program Sedphat,[40] version 8.2, from Dr. Peter Schuck, National Institutes of Health, Bethesda, MD 20892.

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2011.07.063

§ http://mfold.rna.albany.edu/

Acknowledgements

We thank Andy Baron for assistance with the sedimentation velocity experiments and Saskia Bakker for help with Fig. 1, as well as members of the Stockley, Phillips and Twarock research groups for many helpful discussions of the work with the STNV VLP. C.L.L., S.W.L., R.F. and D.H.J.B. were supported by postgraduate studentships from the UK Biotechnology and Biological Sciences Research Council. R.T. and E.C.D. thank the Leverhulme Trust for financial support.

References

1. Schneemann, A. (2006). The structural and functional role of rna in icosahedral virus assembly. *Annu. Rev. Microbiol.* **60**, 51–67.
2. Koning, R., van den Worm, S., Plaisier, J. R., van Duin, J., Pieter Abrahams, J. & Koerten, H. (2003). Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2. *J. Mol. Biol.* **332**, 415–422.
3. Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M. & Johnson, J. E. (2001). The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nat. Struct. Biol.* **8**, 77–83.
4. Toropova, K., Basnak, G., Twarock, R., Stockley, P. G. & Ranson, N. A. (2008). The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J. Mol. Biol.* **375**, 824–836.
5. van den Worm, S. H., Koning, R. I., Warmenhoven, H. J., Koerten, H. K. & van Duin, J. (2006). Cryo electron microscopy reconstructions of the *Leviviridae* unveil the densest icosahedral RNA packing possible. *J. Mol. Biol.* **363**, 858–865.
6. Basnak, G., Morton, V. L., Rolfsson, O., Stonehouse, N. J., Ashcroft, A. E. & Stockley, P. G. (2010). Viral

genomic single-stranded RNA directs the pathway toward a $T=3$ capsid. *J. Mol. Biol.* **395**, 924–936.
7. Dykeman, E. C., Stockley, P. G. & Twarock, R. (2010). Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J. Mol. Biol.* **395**, 916–923.
8. Gell, C., Sabir, T., Westwood, J., Rashid, A., Smith, D. A., Harris, S. A. & Stockley, P. G. (2008). Single-molecule fluorescence resonance energy transfer assays reveal heterogeneous folding ensembles in a simple RNA stem–loop. *J. Mol. Biol.* **384**, 264–278.
9. Lago, H., Fonseca, S. A., Murray, J. B., Stonehouse, N. J. & Stockley, P. G. (1998). Dissecting the key recognition features of the MS2 bacteriophage translational repression complex. *Nucleic Acids Res.* **26**, 1337–1344.
10. Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J. *et al.* (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a $T=3$ viral capsid. *J. Mol. Biol.* **369**, 541–552.
11. Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature*, **371**, 623–626.
12. Golmohammadi, R., Valegard, K., Fridborg, K. & Liljas, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J. Mol. Biol.* **234**, 620–639.
13. Valegard, K., Liljas, L., Fridborg, K. & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature*, **345**, 36–41.
14. Dykeman, E. C., Grayson, N. E., Toropova, K., Ranson, N. A., Stockley, P. G. & Twarock, R. (2011). Simple rules for efficient assembly predict the layout of a packaged viral RNA. *J. Mol. Biol.* **408**, 399–407.
15. Morton, V. L., Dykeman, E. C., Stonehouse, N. J., Ashcroft, A. E., Twarock, R. & Stockley, P. G. (2010). The impact of viral RNA on assembly pathway selection. *J. Mol. Biol.* **401**, 298–308.
16. Rolfsson, O., Toropova, K., Ranson, N. A. & Stockley, P. G. (2010). Mutually-induced conformational switching of RNA and coat protein underpins efficient assembly of a viral capsid. *J. Mol. Biol.* **401**, 309–322.
17. Toropova, K., Stockley, P. G. & Ranson, N. A. (2011). Visualising a viral RNA genome poised for release from its receptor complex. *J. Mol. Biol.* **408**, 408–419.
18. Lane, S. W. (2003). Structural and functional studies of recombinant STNV capsids Ph. D., University of Leeds.
19. Lane, C. L. (2007). Structural studies of protein–RNA interactions in Satellite tobacco necrosis virus (STNV) capsids. Ph. D., University of Leeds.
20. Lane, S. W., Dennis, C. A., Lane, C. L., Trinh, C. H., Rizkallah, P. J., Stockley, P. G. & Phillips, S. E. V. (2011). Construction and crystal structure of recombinant STNV capsids. *J. Mol. Biol.* [Epub ahead of print].
21. Barton, J. L., Bunka, D. H., Knowling, S. E., Lefevre, P., Warren, A. J., Bonifer, C. & Stockley, P. G. (2009). Characterization of RNA aptamers that disrupt the RUNX1-CBFbeta/DNA complex. *Nucleic Acids Res.* **37**, 6818–6830.
22. Bunka, D. H., Mantle, B. J., Morten, I. J., Tennent, G. A., Radford, S. E. & Stockley, P. G. (2007). Production and characterization of RNA aptamers specific for amyloid fibril epitopes. *J. Biol. Chem.* **282**, 34500–34509.
23. Sugiyama, T. & Nakada, D. (1967). Control of translation of MS2 RNA cistrons by MS2 coat protein. *Proc. Natl Acad. Sci. USA*, **57**, 1744–1750.
24. Sorger, P. K., Stockley, P. G. & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly *in vitro*. *J. Mol. Biol.* **191**, 639–658.
25. Unge, T., Montelius, I., Liljas, L. & Ofverstedt, L. G. (1986). The EDTA-treated expanded satellite tobacco necrosis virus: biochemical properties and crystallization. *Virology*, **152**, 207–218.
26. Jones, T. A. & Liljas, L. (1984). Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution. *J. Mol. Biol.* **177**, 735–767.
27. Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lovgren, S., Skoglund, U. & Strandberg, B. (1982). Structure of satellite tobacco necrosis virus at 3.0 Å resolution. *J. Mol. Biol.* **159**, 93–108.
28. Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130.
29. Convery, M. A., Rowsell, S., Stonehouse, N. J., Ellington, A. D., Hirao, I., Murray, J. B. *et al.* (1998). Crystal structure of an RNA aptamer–protein complex at 2.8 Å resolution. *Nat. Struct. Biol.* **5**, 133–139.
30. Hirao, I., Spingola, M., Peabody, D. & Ellington, A. D. (1998). The limits of specificity: an experimental analysis with RNA aptamers to MS2 coat protein variants. *Mol. Diversity*, **4**, 75–89.
31. Rowsell, S., Stonehouse, N. J., Convery, M. A., Adams, C. J., Ellington, A. D., Hirao, I. *et al.* (1998). Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nat. Struct. Biol.* **5**, 970–975.
32. Schneider, D., Tuerk, C. & Gold, L. (1992). Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. *J. Mol. Biol.* **228**, 862–869.
33. Carey, J. & Uhlenbeck, O. C. (1983). Kinetic and thermodynamic characterization of the R17 coat protein–ribonucleic acid interaction. *Biochemistry*, **22**, 2610–2615.
34. Larson, S. B., Day, J., Greenwood, A. & McPherson, A. (1998). Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *J. Mol. Biol.* **277**, 37–59.
35. Larson, S. B., Koszelak, S., Day, J., Greenwood, A., Dodds, J. A. & McPherson, A. (1993). Double-helical RNA in satellite tobacco mosaic virus. *Nature*, **361**, 179–182.
36. Larson, S. B., Koszelak, S., Day, J., Greenwood, A., Dodds, J. A. & McPherson, A. (1993). Three-dimensional structure of satellite tobacco mosaic virus at 2.9 Å resolution. *J. Mol. Biol.* **231**, 375–391.
37. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415.
38. Durchschlag, H. (1986). In *Thermodynamics Data for Biochemistry and Biotechnology* (Hinz, H. J., ed.), Springer Verlag, New York, NY.
39. Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619.
40. Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A. & Schuck, P. (2004). Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal. Biochem.* **326**, 234–256.